

Characterization and Computation of Nash Equilibria in Horizontal Queueing Networks

Walid Krichene

Jack Reilly

Saurabh Amin

Alexandre M. Bayen

Abstract—We study inefficiencies in horizontal queueing networks due to the selfish behavior of agents, by comparing social optima to Nash equilibria. The article expands studies on routing games which traditionally model congestion with latency functions that increase with the flow on a particular link. This type of latency function cannot capture congestion effects on horizontal queueing networks. Latencies on horizontal queueing networks increase as a function of density, and flow can decrease with increasing latencies. For example, this class of latency functions arises in transportation networks. For static analysis of horizontal queueing on parallel-link networks, we show that there are multiple Nash equilibria with different total costs, which contrasts results for increasing latency functions. We present a novel algorithm, quadratic in the number of links, for computing the Nash equilibrium that minimizes total cost (best Nash equilibrium). The relative inefficiencies of best Nash equilibria are evaluated through analysis of the price of stability, and analytical results are presented for two-link networks. Price of stability is shown to be sensitive to changes in demand when links near capacity, and congestion mitigation strategies are discussed, motivated by our results.

I. INTRODUCTION

A. Routing games and Nash equilibria

Routing games (or congestion games) form an important class of non-atomic games that is used to model the interaction of agents who are using resources on a network, in which the cost on each edge depends on the fraction of agents using that edge. Extensive work has been dedicated to studying Nash equilibria (or user optimal assignments) of congestion games [9], [11], [14], in which all players are assumed to choose the routes that minimize their respective individual costs. Under some assumptions on the latency functions, Nash equilibria can be computed as a solution of a convex optimization problem [6]. Nash equilibria of congestion games are known to be inefficient compared to system optimal assignments, in which a coordinator, or a central authority, assigns flow as to minimize a cost function over all players [2], [17]. Other variants of congestion games exist in the game theory literature [3], [10].

B. A new class of latency functions

The class of latency functions that have been studied so far in routing games literature rely on the following assumptions: if $l(x)$ is the latency on a link, where x is the flow, then l is assumed to be non-decreasing, and $x \mapsto xl(x)$ is assumed to be convex [13]. While this class of latency functions provides a good model of congestion for a considerable range of networks, such as communication networks, it does not accurately model horizontal queueing congestion, such as congestion on transportation networks [4], [16], [8].

Intuitively, a given flow x on a road segment can correspond to

- either a large concentration of agents moving slowly (high density on a congested road), in which case the latency is large,
- or few cars moving quickly (low density), in which case the latency is small.

Due to this phenomenon, we later show that the latency is not uniquely determined by the flow, and depends on the congestion state of the link. One simple way of modeling this phenomenon is to have an additional binary argument m in the latency function $l(x, m)$ to specify whether the link is congested ($m = 1$) or is in free-flow ($m = 0$). For example, in transportation networks, latency functions derived from a macroscopic model of traffic flow developed by Lighthill and Whitham [8], can be expressed in the above form $l(x, m)$. One interesting result is that the latency under congestion $l(x, 1)$ is a *decreasing* function of flow. Intuitively, as the link becomes more congested, agents slow down, so their latency increases, and the amount of flow on the link decreases.

A large body of literature in horizontal queueing theory has studied game-theoretic concepts, such as dynamic traffic assignment for user equilibria [5], [9] and system optimal assignments [17]. Due to the complexities of modeling horizontal queueing, approaches to solving the user equilibrium on general networks usually involve non-linear optimization techniques that limit the size of networks that can be considered. By limiting our analysis to parallel-link networks, we exploit the structure of the network to improve upon previous approaches to user equilibria computation.

C. Contributions of the article

We introduce a new class of latency functions that is expressive enough to model congestion on horizontal queueing networks, and study, for this new class of latency functions, the Nash equilibria of the congestion game on parallel networks. We consider a congestion game on a parallel network, where each link has a dual-mode latency function: latency is constant in free-flow, and a decreasing function of flow in congestion. This leads to novel theoretical and algorithmic results for characterizing and computing Nash equilibria:

- We show that there is no essential uniqueness of Nash equilibria (not all Nash equilibria have equal total costs), unlike point-queueing models usually considered in congestion games [14].
- We show that for a given instance (N, r) of a parallel network of N links, subject to a constant demand r , we

characterize the structure of the flow in the *best Nash equilibrium* (the Nash equilibrium that minimizes the total network latency) and show that the equilibrium can be computed in $O(N^2)$ time.

- We give analytical solution to the price of stability on a two-link parallel network. This gives insight into the qualitative behavior of congestion caused by Nash equilibria on horizontal queueing networks. In particular, when the lowest-latency link in a network nears capacity, diverting only a small amount of flow to a slower link can avert congestion completely.

These results provide a framework for efficient computation of Nash equilibria on parallel link networks, which, in turn, give a high-level explanation of congestion patterns on such networks. While the assumption of parallel-link networks may seem an overly-limiting assumption, there are many examples of highway networks in populous areas (such as the California bay area), in which such networks can model congestion via horizontal queueing and the parallel link structure is an accurate description. Additionally, highway networks often suffer from congestion due to selfish routing, and would benefit from the analysis of horizontal queueing Nash equilibria.

D. Organization

We start by defining the model and introducing a new class of latency functions in Section II, and show as an example how such latency functions can be derived from known macroscopic models of traffic flow. In Section III, we study Nash equilibria of congestion games for this new class of latency, and show that the essential uniqueness property does not hold. We then bound the number of Nash equilibria and give a tractable algorithm for computing the set of Nash equilibria. In Section IV we characterize in particular the best Nash equilibrium and give an explicit algorithm for its computation. In Section V-B, we study the inefficiency of best Nash equilibria using price of stability as the measure. We conclude with a summary of our results and directions for future work in Section VI.

II. THE MODEL

A. Congestion Games on Parallel Edge Network

We consider a non-atomic¹ congestion game on a parallel network similar to the one studied in [13], shown in Figure 1. The network has a single source and a single sink. Connecting the source and sink are N parallel edges (or links) indexed by $n \in \{1, \dots, N\}$. The network is subject to a constant positive flow demand r at the source. We will denote by (N, r) an instance of a network with N parallel links subject to demand r . A feasible flow assignment for the instance (N, r) is a vector $x \in \mathcal{R}_+^N$ such that $\sum_{n=1}^N x_n = r$ where x_n is the flow on link n .

¹When fractional flows are allowed, the players are said to be non-atomic [15]. The choice of atomic versus non-atomic players in congestion games is similar to the modeling choice of microscopic versus macroscopic flow framework in traffic modeling [8].

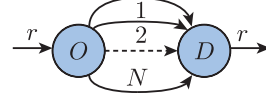


Fig. 1: Network instance (N, r) . N links from O to D with demand r .

We then introduce a cost function, or latency function l_n , on each link n . A link's cost can be thought of as the latency experienced by a job assigned to a particular machine n in the case of job scheduling [13], or the travel time of a vehicle using a particular road n in the case of traffic assignment.

B. Modeling congestion with Latency Functions

To model the effects of queueing on a given link n , the latency l_n on the link is typically assumed to be a non-decreasing function of the amount of flow x_n on link n [2], [3], [14]. While this class of latency accurately models congestion on a broad range of networks (such as the internet, and more generally communication networks), it fails to correctly model congestion for a large class of networks. For example, consider a link (or road) n in a traffic network. A given flow x_n may correspond to two different scenarios: few vehicles on the road are moving quickly (the road is in *free-flow*), in which case the latency on the road is low, or a large number of vehicles on the road are moving slowly (the road is *congested*), in which case the latency on the road is high. This phenomenon is not captured if the latency is only a function of flow, as such functions do not allow capacity to decrease as latencies increase (Figure 2c shows how mapping flow to latency is not unique). One way to address this limitation is to have two latency functions: one describes the latency on a link in *free-flow*, while the other describes the latency on a link in *congestion*. Equivalently, one may introduce a second binary argument $m_n \in \{0, 1\}$ to the latency function, designed to specify whether the link is in free-flow or in congestion.

We next show that such latency functions can be derived from macroscopic models of flow on Horizontal Queueing networks.

C. Deriving Latency Functions for Horizontal Queueing Networks

The relationship between the speed of flow on a network and the *density* of flow (or amount of flow in the static sense) is usually expressed by a function called the flux function in the physical sciences and conservation law theory and fundamental diagram in traffic flow theory [4], [12]. Figure 2a depicts a triangular flux function, while similarly shaped diagrams have been developed for certain applications.

While such flow models have been popular for many decades in specific domains (such as traffic and fluid mechanics), less attention has been given to these models in the literature studying congestion games, which focuses on modeling latency as a non-decreasing function of flow, and assumes flow and density to have a one-to-one relation. In order to characterize Nash equilibria on horizontal queueing

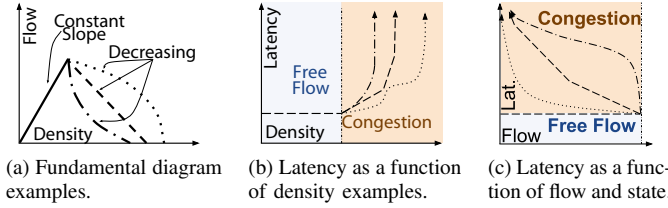


Fig. 2: Examples of flux functions and derived relationships that satisfy the assumptions in Section II-C. The dotted line describes a flux function which reacts less severely to increasing density than the dashed and dot-dashed lines.

networks, we develop a novel approach based on the unique structure of the latency functions.

Consider a link n with length L_n , and assume the flow x_n on the link is given by a function:

$$\begin{aligned} x_n^\rho : D_n &\rightarrow \mathcal{R}_+ \\ \rho_n &\mapsto x_n = x_n^\rho(\rho_n) \end{aligned}$$

of a given density ρ_n on the link. The domain of ρ_n , D_n , is non-negative and potentially unbounded. The latency is given by a function:

$$\begin{aligned} l_n^\rho : D_n &\rightarrow \mathcal{R}_+ \\ \rho_n &\mapsto l_n^\rho(\rho_n) \end{aligned}$$

The function x_n^ρ maps density to flow, and is defined on the domain $D_n \subset \mathcal{R}_+$, and corresponds to the fundamental diagram of traffic². First, we observe that latency is related to flow and density through the relation:

$$l_n^\rho(\rho_n) = \frac{L_n \rho_n}{x_n^\rho(\rho_n)}, \quad (1)$$

We make three assumptions on the flow and latency functions x_n^ρ and l_n^ρ for horizontal queues, which are illustrated in Figure 2a.

- 1) There exists a critical density $\rho_n^{\text{crit}} > 0$, such that the latency is constant and minimal on the interval $[0, \rho_n^{\text{crit}}]$. This is equivalent to the flow function x_n^ρ being linear on that interval.
- 2) x_n^ρ is maximal at ρ_n^{crit} . The value of the flow at critical density is denoted $x_n^{\text{max}} = x_n^\rho(\rho_n^{\text{crit}})$ and referred to as maximum flow or *capacity* of the link.
- 3) The latency function l_n^ρ , is continuous, non-decreasing.

We define the *free-flow region* as $\rho_n \in [0, \rho_n^{\text{crit}}]$ and *congested region* as $\rho_n > \rho_n^{\text{crit}}$. These assumptions define a class of supported fundamental diagrams. Assumption 2 just states that $\forall \rho_n \geq 0$, $x_n^\rho(\rho_n) \in [0, x_n^{\text{max}}]$. From Equation (1), Assumption 3 can be expressed equivalently in terms of the flow function x_n^ρ as $\frac{dx_n^\rho(\rho_n)}{d\rho_n} < \frac{x_n^\rho(\rho_n)}{\rho_n}$. This gives reasonable restrictions on the shape of fundamental diagrams in the congestion region, and flexible enough to allow non-concavity, and even increasing functions x_n^ρ under certain conditions. Examples of allowable fundamental diagrams

are given in Figure 2a, and corresponding examples of latency functions are given in Figure 2b. Note that from these assumptions, we can write the latency function for the horizontal queueing model as:

$$l_n^\rho(\rho_n) = \begin{cases} \frac{L_n \rho_n^{\text{crit}}}{x_n^{\text{max}}} & \rho_n \in [0, \rho_n^{\text{crit}}] \\ \frac{L_n \rho_n}{x_n(\rho_n)} & \text{otherwise} \end{cases}$$

Example 1: Triangular fundamental diagrams

One particular class of fundamental diagrams x^ρ that satisfy the above assumptions are triangular fundamental diagrams [4], which are linear with positive slope v^f in the free-flow region, affine with negative slope v^c in the congestion region, and have maximum flow $x^{\text{max}} = \rho^{\text{crit}} v^f$. Assumptions 1 and 2 are satisfied by definition, and Assumption 3 is satisfied since $\frac{dx(\rho)}{d\rho} = v^f = \frac{x(\rho)}{\rho} \forall \rho \in [0, \rho^{\text{crit}}]$ and $\frac{dx(\rho)}{d\rho} = v^c \leq 0 \leq \frac{x(\rho)}{\rho} \forall \rho \geq \rho^{\text{crit}}$. The dotted line in Figure 2a shows a triangular fundamental diagram. The latency function is then given by:

$$l_\Delta^\rho(\rho) = \begin{cases} \frac{L}{v^f} & 0 \leq \rho \leq \frac{x^{\text{max}}}{v^f} \\ \frac{L\rho}{v^c(\rho - \rho^{\text{max}})} & \frac{x^{\text{max}}}{v^f} < \rho \leq \rho^{\text{max}} \end{cases}$$

where $\rho^{\text{max}} = x^{\text{max}} \left(\frac{1}{v^f} - \frac{1}{v^c} \right)$.

D. A new class of Latency functions: Horizontal Queueing Latency

While expressing latency as a function of density is intuitive and succinct for horizontal queues, expressing it as a function of flow proves to be more convenient in the study of congestion games. This is largely due to the fact that total flow must be conserved in traffic assignment problems, and not density. For this reason, we introduce an equivalent formulation of latency using flow and a binary argument that describes congestion state. Let the congestion state m_n of link n be defined as:

$$m_n := \begin{cases} 0 & \text{if } n \text{ is in free-flow} \\ 1 & \text{if } n \text{ is congested} \end{cases}$$

We can now define a general class of latency functions l_n as a function of both flow and congestion state:

$$\begin{aligned} l_n : [0, x_n^{\text{max}}] \times \{0\} \cup (0, x_n^{\text{max}}] \times \{1\} &\rightarrow \mathcal{R}_+ \\ (x_n, m_n) &\mapsto l_n(x_n, m_n), \end{aligned}$$

Note that the latency in congestion $l_n(\cdot, 1)$ is defined on the open interval $(0, x_n^{\text{max}}]$. In particular, if $x_n = 0$ then $m_n = 0$ (a link that is not in the support is considered in free-flow) and if $x_n = x_n^{\text{max}}$ then $m_n = 0$ (if a link is at maximum capacity, it is considered, by convention, to be in free-flow. It is in fact on the boundary of the free-flow and congestion regions, and we choose this convention for simplicity of notation). We also assume that l_n satisfies the following properties, which are equivalent to the assumptions in Section II-C:

- The latency in free-flow is constant. Equivalently, $\forall x_n \in [0, x_n^{\text{max}}]$, $l_n(x_n, 0) = a_n$, where a_n is the constant free-flow latency.
- $\lim_{x_n \rightarrow x_n^{\text{max}}} l_n(x_n, 1) = l_n(x_n^{\text{max}}, 0) = a_n$

²Note, this allows for fundamental diagrams with unbounded density support, for example as in [12]: $x(\rho) = \rho e^{-\rho}$

- $x \mapsto l_n(x, 1)$ is decreasing on $(0, x_n^{\max})$.

Some examples of latency functions in this class are illustrated in Figure 2c. Again, the latency function corresponding to a triangular fundamental diagram can be readily expressed in this form:

$$\begin{aligned} l_{\Delta}(x, 0) &= \frac{L}{vf} \\ l_{\Delta}(x, 1) &= L \left(\frac{\rho^{\max}}{x} + \frac{1}{v^c} \right) \end{aligned}$$

E. Total System Cost

The cost to a single user is defined as the latency experienced by the user, or the latency of the link chosen by the user. Therefore, the total cost experienced on a particular link $C_n(x_n, m_n) = l_n(x_n, m_n) x_n = L_n \rho_n$. Then, the total system cost is the sum of the costs of the individual links $C(x, m) = \sum_{i=1}^N C_i(x_i, m_i)$, where $x = (x_1, \dots, x_N)$ is the vector of flows, and $m = (m_1, \dots, m_N)$ is the vector of congestion states for the entire network.

III. NASH EQUILIBRIA

In this section, we characterize pure non-atomic Nash equilibria of the network (also called Wardrop equilibria in the transportation literature), which we simply refer to as Nash equilibria. We show that our class of latency functions induce multiple Nash equilibria with different costs, and that the set of Nash equilibria can be computed in polynomial time (with respect to the number of parallel links). Then we characterize the best Nash equilibrium and focus our attention on studying the inefficiency of the best Nash equilibrium in Section V.

A. Characterization of Nash Equilibria

We first recall the fundamental notion of Nash equilibrium for the network instance (N, r) .

Definition 1: Nash Equilibrium

An assignment $(x, m) \in \mathcal{R}_+^N \times \{0, 1\}^N$ for the network instance (N, r) is a Nash equilibrium, if $\forall n$

$$x_n > 0 \Rightarrow \forall k \leq N, l_n(x_n, m_n) \leq l_k(x_k, m_k)$$

In particular, every non-atomic agent cannot improve her/his latency by switching to another link. As a consequence, all links that are in the support of x have the same latency l_0 , and links that are not in the support have latency greater than or equal to l_0 . We will denote by $\text{Supp}(x)$ the support of x , i.e. the set $\{n \in \{1, \dots, N\} | x_n > 0\}$.

Note that to fully describe the equilibrium, one needs to specify the congestion state vector m in addition to the flow assignment x , since the latency on a link depends on whether the link is congested or not. The following Lemma gives an equivalent characterization of Nash equilibria.

Lemma 1: Characterization of a Nash Equilibrium

A feasible assignment (x, m) for a network instance (N, r) is a Nash equilibrium if and only if $\exists l_0 > 0$ such that

$$\begin{aligned} x_n > 0 &\Rightarrow l_n(x_n, m_n) = l_0 \\ x_n = 0 &\Rightarrow l_n(0, 0) \geq l_0 \end{aligned}$$

The total latency incurred by the network is $C(x, m) = r l_0$. Note that links that have zero flow are necessarily in free-flow $x_n = 0 \Rightarrow m_n = 0$.

B. Horizontal queueing networks have multiple Nash equilibria

Let $\text{NE}(N, r)$ denote the set of Nash Equilibria for network instance (N, r) . For our class of latency functions, the essential uniqueness property of Nash equilibrium [14] does not hold.³ To see this, consider for example a network instance $(N=2, r=1)$ where $x_1^{\max} = x_2^{\max} = 1$, the free-flow latencies are $a_1 = 1$ and $a_2 = 2$, and the congested latency functions are given respectively by $l_1(x_1, 1) = \frac{1}{x_1}$ and $l_2(x_2, 1) = \frac{2}{x_2}$. Then it is easy to see that $(x, m) = ((1, 0), (0, 0))$, $(x', m') = ((\frac{1}{2}, \frac{1}{2}), (1, 0))$, and $(x'', m'') = ((\frac{1}{3}, \frac{2}{3}), (1, 1))$ are all Nash equilibria for this instance, and they have different costs: $C(x, m) = 1$, $C(x', m') = 2$ and $C(x'', m'') = 3$. This simple example shows that there are at least two types of Nash equilibria: equilibria for which every link in the support is congested (this is the case for (x'', m'') in the previous example), and equilibria that have one link in free-flow in their support (this is the case for both (x, m) and (x', m')). In this section, we show that these are in fact the only possible types of equilibria, and we prove that there are at most $2N$ such equilibria, assuming free-flow latencies are distinct. To simplify the discussion, we assume without loss of generality, that the links are ordered by increasing free-flow latencies, and that free-flow latencies are different to avoid degenerate cases where the set of Nash equilibria is infinite ($a_1 < a_2 < \dots < a_N$).

We start by deriving properties that the congestion state vector m needs to satisfy for a Nash equilibrium (x, m) . We first show that if a link j is in the support of x , then links $\{1, \dots, j-1\}$ are congested.

Lemma 2: Congestion of lower links

Let $(x, m) \in \text{NE}(N, r)$.

Then $j \in \text{Supp}(x) \Rightarrow m_i = 1 \quad \forall i \in \{1, \dots, j-1\}$

Proof: Let $i \in \{1, \dots, j-1\}$. Then $m_i = 0 \Rightarrow l_i(x_i, m_i) = a_i < a_j \leq l_j(x_j, m_j)$, which violates the characterization of Nash equilibrium in Lemma 1. Therefore, $m_i = 1 \quad \forall i \in \{1, \dots, j-1\}$. ■

Corollary 1: Congestion states under equilibrium

Let $(x, m) \in \text{NE}(N, r)$. Assume that $\exists j \in \text{Supp}(x)$ such that $m_j = 0$. Then $m = (1, \dots, 1, 0, \dots, 0)$ and $\text{Supp}(x) = \{1, \dots, j\}$.

Proof: We have from Lemma 2 that $\forall i \in \{1, \dots, j-1\}$, $m_i = 1$. And we have $\forall i \in \{j+1, \dots, N\}$,

³Note that essential uniqueness of Nash Equilibria holds for the class of non decreasing latency functions, i.e. all Nash Equilibria have the same cost. To show this result, assume that the latency functions $x_n \mapsto l_n(x_n)$ are non-decreasing and only depend on the flow x_n . Let x and x' be two Nash equilibria for (N, r) . Let l_0 , respectively l'_0 denote the common latency of all links in the support of x , respectively x' . The cost of the Nash equilibria are respectively $r l_0$ and $r l'_0$. Assume $x \neq x'$. Then $\exists n_1, n_2$ such that $x_{n_1} > x'_{n_1} \geq 0$ and $x'_{n_2} > x_{n_2} \geq 0$. Since x is at Nash equilibrium and $n_1 \in \text{Supp}(x)$, $l_{n_1}(x_{n_1}) \leq l_{n_2}(x_{n_2})$. And since l_{n_2} is non-decreasing $l_{n_2}(x_{n_2}) \leq l_{n_2}(x'_{n_2})$. Thus $l_0 = l_{n_1}(x_{n_1}) \leq l_{n_2}(x_{n_2}) \leq l_{n_2}(x'_{n_2}) = l'_0$. Exchanging the roles of x and x' we have $l'_0 \leq l_0$. Therefore $l_0 = l'_0$ and both equilibria have the same cost.

$l_i(x_i, m_i) \geq a_i$ by definition of the latency function, and $a_i > a_j$ since $i > j$. Therefore the latency on link $i \in \{j+1, \dots, N\}$ is strictly greater than the latency on link $j \in \text{Supp}(x)$, therefore $i \notin \text{Supp}(x)$ (follows from the characterization of Nash equilibrium in Lemma 1) and $m_i = 0$. ■

The corollary states that if some link k in the support of a Nash equilibrium is in free-flow, this completely determines the congestion state vector of the equilibrium: links $\{1, \dots, k-1\}$ are in the support and are congested, and links $\{k+1, \dots, N\}$ are not in the support. We will call such Nash equilibria (where a single link in the support is in free-flow) *single link free-flow equilibria*. In general a Nash equilibrium does not necessarily have a link in free-flow: this defines a second type of equilibria where all links in the support are congested, i.e. $m_{\max \text{Supp}(x)} = 1$. We will call such equilibria *congested equilibria*.

The following Lemma shows that given a congestion state vector m , there are at most two corresponding Nash equilibria (x, m) , one single link free-flow equilibrium, and one congested equilibrium.

Lemma 3: Enumerating Nash Equilibria

For a given congestion state m , there are at most two assignments x such that (x, m) is a Nash equilibrium.

Proof: Let $m \in \{0, 1\}^N$ be a given congestion state vector and assume $x, x' \in \mathcal{R}_+^N$ are such that (x, m) and (x', m) are two different Nash equilibria. Then $\exists i, j, 1 \leq i < j \leq N$ such that $0 \leq x_i < x'_i$ and $0 \leq x'_j < x_j$ (we assume without loss of generality that $i < j$: if this is not the case, exchange x and x').

We start by noting that since $j \in \text{Supp}(x)$ and $i < j$, then $i \in \text{Supp}(x)$. This follows from the fact that $l_i(0, m_i) = a_i < a_j \leq l_j(x_j, m_j)$, thus if $j \in \text{Supp}(x)$, x_i cannot be zero since every link in the support of a Nash equilibrium has latency \leq the latency on any other link.

Now since $i, j \in \text{Supp}(x)$, then we have $l_i(x_i, m_i) = l_j(x_j, m_j)$. And since $j \in \text{Supp}(x)$ and $i < j$, then by Lemma 2, we have $m_i = 1$ (link i is congested). Therefore $l_i(x_i, m_i) > l_i(x'_i, m_i)$ since $l_i(\cdot, 1)$ is decreasing. Finally we have $l_j(x_j, m_j) \leq l_j(x'_j, m_j)$ since $l_j(\cdot, 0)$ is constant and $l_j(\cdot, 1)$ is decreasing. Combining the above, we have

$$l_i(x'_i, m_i) < l_i(x_i, m_i) = l_j(x_j, m_j) \leq l_j(x'_j, m_j) \quad (2)$$

Now we partition the set of Nash Equilibria in two sets $\text{NE}_=(N, r) \sqcup \text{NE}_f(N, r) \sqcup \text{NE}_c(N, r)$: equilibria that have a completely congested support, denoted by $\text{NE}_c(N, r)$, and equilibria that have one link in free-flow in their support, denoted by $\text{NE}_f(N, r)$.

$$\begin{aligned} \text{NE}_f(N, r) &= \{(x, m) \in \text{NE}_l(N, r) \mid m_{\max \text{Supp}(x)} = 0\} \\ \text{NE}_c(N, r) &= \{(x, m) \in \text{NE}_l(N, r) \mid m_{\max \text{Supp}(x)} = 1\} \end{aligned}$$

Now we show that for a given congestion state vector m , each set contains at most one element.

Suppose $(x, m), (x', m) \in \text{NE}_f(N, r)$, where x, x' are as defined above. Then since $j \in \text{Supp}(x)$, we have by Lemma 2, $\forall k < j$ $m_k = 1$. Since the last link in the support of x' is, by assumption, in free-flow, we have

$\max \text{Supp}(x') \geq j$. Therefore $j \in \text{Supp}(x')$. But from (2) we have $l_i(x'_i, m_i) < l_j(x'_j, m_j)$ which contradicts the definition of a Nash Equilibrium (a link in the support of a Nash Equilibrium has latency less than or equal to any other link). Thus there is at most one assignment x such that $(x, m) \in \text{NE}_f(N, r)$.

Suppose $(x, m), (x', m) \in \text{NE}_c(N, r)$, where x, x' are as defined above. Then since $j \in \text{Supp}(x)$ and every link in the support is congested (by definition of $\text{NE}_c(N, r)$), then $m_j = 1$. Therefore j is also congested under assignment x' , thus $j \in \text{Supp}(x')$. Similarly to the first case, this leads to a contradiction since $l_i(x'_i, m_i) < l_j(x'_j, m_j)$, which proves that there is at most one assignment x such that $(x, m) \in \text{NE}_c(N, r)$. ■

This shows that there are at most $2N$ Nash equilibria for the instance (N, r) : N single link free-flow equilibria, corresponding to congestion states $m = (0, \dots, 0)$, $m = (1, 0, \dots, 0)$, \dots , $m = (1, \dots, 1, 0)$, and N congested equilibria, corresponding to congestion states $m = (1, 0, \dots, 0)$, \dots , $m = (1, \dots, 1)$. Next, we characterize single link free-flow equilibria.

C. Single link free-flow Equilibria

Consider a Nash equilibrium (x, m) and let $k = \max \{\text{Supp}(x)\}$. Assume $m_k = 0$ (i.e. (x, m) is a free-flow Nash equilibrium). We have from Corollary 1 that links $\{1, \dots, k-1\}$ are congested and link k is in free-flow. Therefore we must have $\forall n \in \{1, \dots, k-1\}$, $l_n(x_n, 1) = l_k(x_k, 0) = a_k$. This uniquely determines the flow on congested links $n \in \{1, \dots, k-1\}$. We define this flow to be $\hat{x}_n(k)$. More precisely,

Definition 2: Congestion flow

For $1 \leq n < k \leq N$, the congestion flow $\hat{x}_n(k)$ is defined as the unique flow in $(0, x_n^{\max})$ that satisfies

$$l_n(\hat{x}_n(k), 1) = a_k \quad (3)$$

Proposition 1: Congestion Flows are decreasing

$$\hat{x}_n(k) = l_n(\cdot, 1)^{-1}(a_k) \quad (4)$$

is a decreasing function of k since a_k is increasing in k and $l_n(\cdot, 1)^{-1}$ is decreasing.

We can now characterize single link free-flow equilibria. All single link free-flow equilibria are of the form $(\bar{x}^{k,r}, \bar{m}^k)$ where

$$\bar{m}^k := (1, \dots, 1, 0, \dots, 0) \quad (5)$$

$$\bar{x}^{k,r} := (\hat{x}_1(k), \dots, \hat{x}_{k-1}(k), r - \sum_{n=1}^{k-1} \hat{x}_n(k), 0, \dots, 0) \quad (6)$$

Illustrations of Equations (3), (5) and (6) are shown in Figure 3.

Proposition 2: Single link free-flow Nash Equilibria

If $\bar{x}^{k,r}$ is a feasible assignment, i.e. $r - \sum_{n=1}^{k-1} \hat{x}_n(k) \in [0, x_k^{\max}]$, then $(\bar{x}^{k,r}, \bar{m}^k)$ is a Nash Equilibrium for the instance (N, r) .

Proof: From (5) and (6), we have that $\forall i < k$, $\bar{x}_i^{k,r} \in [0, x_i^{\max}]$ and $l_i(\bar{x}_i^{k,r}, \bar{m}_i^k) = a_k$. And since $\bar{m}_k^k = 0$ by definition of \bar{m}_k^k , we also have $l_k(\bar{x}_k^{k,r}, \bar{m}_k^k) = a_k$. All links $n > k$ are not in $\text{Supp}(\bar{x}^{k,r})$, and have a latency

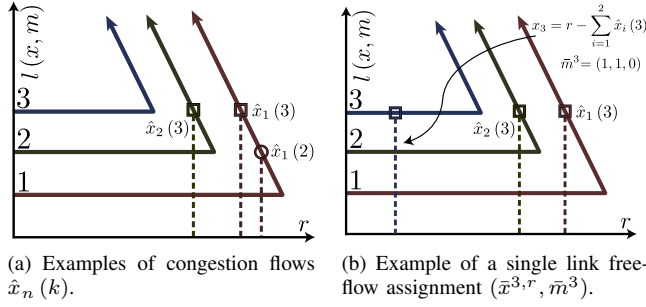


Fig. 3: Graphical illustration of single link free-flow Nash equilibria.

greater than a_k . Therefore, we have that $\forall n \in \text{Supp}(\bar{x}^{k,r})$, $l_n(\bar{x}_n^{k,r}, \bar{m}_n^k) = a_k$ and $\forall n \notin \text{Supp}(\bar{x}^{k,r})$, $l_n(\bar{x}_n^{k,r}, \bar{m}_n^k) > a_k$, which satisfies the definition of a Nash equilibrium. ■

D. Existence of a single-link free-flow Nash Equilibrium

From property 2, we have a simple characterization of single link free-flow equilibria. Next, we show that if the set of Nash equilibria is non-empty, then it contains a single link free-flow equilibrium.

Lemma 4: Existence of a single link free-flow Nash equilibrium

Consider instance (N, r) . If the set of Nash equilibria is non empty, $\text{NE}(N, r) \neq \emptyset$, then there exists a single link free-flow Nash equilibrium $(\bar{x}^{j,r}, \bar{m}^j) \in \text{NE}(N, r)$ for some $j \leq N$.

Proof: We first observe that for a network of N links, the maximum demand r such that $\text{NE}(N, r) \neq \emptyset$ is $\max_{k \in \{1, \dots, N\}} \{x_k^{\max} + \sum_{n=1}^{k-1} \hat{x}_n(k)\}$. We denote this quantity with $r^{\text{NE}}(N)$. Therefore, from Lemma 2, it suffices to show the following property:

\mathbf{P}_N : $\forall r \in [0, r^{\text{NE}}(N)]$, there exists a single link free-flow Nash equilibrium for the instance (N, r) .

We show \mathbf{P}_N by induction on N , the size of the network. For $N = 1$, it is clear that if $0 \leq r \leq x_1^{\max}$, there is a single-link free-flow equilibrium $(x, m) = (r, 0)$.

Now let $N \geq 1$, assume \mathbf{P}_N is true and let us show \mathbf{P}_{N+1} . Let $0 \leq r \leq r^{\text{NE}}(N+1)$ and consider a network instance $(N+1, r)$.

Case 1: If $r \leq r^{\text{NE}}(N)$, then by the induction hypothesis there exists a single link free-flow Nash equilibrium (x, m) for the instance (N, r) . Then assignment (x', m') defined as $x' = (x_1, \dots, x_N, 0)$ and $m' = (m_1, \dots, m_N, 0)$ is clearly a single-link free-flow Nash equilibrium for the instance $(N+1, r)$.

Case 2: If $r^{\text{NE}}(N) < r \leq r^{\text{NE}}(N+1)$ then we can show that $(\bar{x}^{N+1,r}, \bar{m}^{N+1}) \in \text{NE}(N+1, r)$. From Proposition 2, we only need to show that

$$0 \leq r - \sum_{n=1}^N \hat{x}_n(N+1) \leq x_{N+1}^{\max}. \quad (7)$$

First, we note that since $r^{\text{NE}}(N) < r^{\text{NE}}(N+1)$, then $r^{\text{NE}}(N+1) = x_{N+1}^{\max} + \sum_{n=1}^N \hat{x}_n(N+1)$, thus from $r < r^{\text{NE}}(N+1)$, we have $r \leq x_{N+1}^{\max} + \sum_{n=1}^N \hat{x}_n(N+1)$

which proves the second inequality of (7). To show the first inequality, we have

$$\begin{aligned} r &\geq x_N^{\max} + \sum_{n=1}^{N-1} \hat{x}_n(N) && \text{since } r^{\text{NE}}(N) < r \\ &\geq x_N^{\max} + \sum_{n=1}^{N-1} \hat{x}_n(N+1) && \text{since } \hat{x}_n(N) \geq \hat{x}_n(N+1) \\ &\geq \hat{x}_N(N+1) + \sum_{n=1}^{N-1} \hat{x}_n(N+1) && \text{since } x_N^{\max} \geq \hat{x}_N(N+1) \end{aligned}$$

which achieves the induction. ■

Corollary 2: Cost of Equilibria

If there exists a congested equilibrium $(x, m) \in \text{NE}(N, r)$, then there exists a single-link free-flow equilibrium (x', m') with lower cost.

Proof: Let $(x, m) \in \text{NE}(N, r)$ be a congested equilibrium, i.e. $m_k = 1$ where $k = \max \text{Supp}(x)$. Then we have $r \leq x_k^{\max} + \sum_{n=1}^{k-1} \hat{x}_n(k)$ and by the property \mathbf{P}_k , there exists a single-link free-flow equilibrium $(\tilde{x}, \tilde{m}) \in \text{NE}(k, r)$, and the cost of this equilibrium is $C(\tilde{x}, \tilde{m}) \leq a_k r$. But this also provides a single-link free-flow equilibrium (x', m') for the original instance (N, r) defined by $x' = (\tilde{x}_1, \dots, \tilde{x}_k, 0, \dots, 0)$ and $m' = (\tilde{m}_1, \dots, \tilde{m}_k, 0, \dots, 0)$, and $C(x', m') = C(\tilde{x}, \tilde{m}) \leq a_k r$. To conclude, we simply note that the cost of the congested equilibrium is $C(x, m) = l_k(x_k, 1)r > a_k r$, thus $C(x, m) > C(x', m')$. ■

IV. BEST NASH EQUILIBRIA

A. Determining minimum cost Nash equilibria

In order to study the inefficiency of Nash equilibria, we focus our attention on *best Nash equilibria* and *price of stability* as a measure of their inefficiency (see for example [1]). A *best Nash equilibrium* (BNE) is defined to be a Nash equilibrium of least total latency.

Definition 3: Best Nash Equilibrium

$$\text{BNE}(N, r) = \arg \min_{(x, m) \in \text{NE}(N, r)} C(x, m)$$

We now show some interesting properties of the best Nash equilibrium:

- 1) BNE (N, r) is unique.
- 2) BNE (N, r) is a single-link free-flow equilibrium.
- 3) BNE (N, r) has the smallest support of all Nash equilibria for demand r .

These properties are summarized in the following theorem.

Theorem 1: Characterization of Best Nash Equilibria

For a parallel network instance (N, r) , the unique best Nash equilibrium is the single-link free-flow equilibrium that has smallest support:

$$\text{BNE}(N, r) = \arg \min_{(x, m) \in \text{NE}(N, r)} \{\max[\text{Supp}(x)]\}$$

Proof: From Corollary 2 we have that if $(x, m) \in \text{NE}(N, r)$ is a congested equilibrium, then there exists a single-link free-flow equilibrium with lower cost. Therefore the Best Nash Equilibrium is a single-link free-flow equilibrium. To show that the BNE has smallest support, we

simply note that if $(x, m) \in \text{NE}_f(N, r)$ is a single-link free-flow equilibrium and $k = \max \text{Supp}(x)$, then its cost is $C(x, m) = a_k r$. Note that uniqueness is immediate since two single-link free-flow equilibria (x, m) and (x', m') that have the same support, hence the same congestion state $m = m'$, coincide by Lemma 3. ■

Theorem 1 therefore provides a simple characterization of the best Nash equilibrium for any instance (N, r) . This characterization results in a simple algorithm to compute the best Nash equilibrium for any network and any feasible demand.

B. Computational complexity of finding Best Nash Equilibria

In this section, we present a constructive algorithm for finding the best Nash equilibrium of a network instance (N, r) and then show the running time to be in $O(N^2)$.

Algorithm (1) relies on the routine `freeFlowConfig`, which outputs a candidate *single-link-free-flow* assignment for the instance (N, r) , such that link i is the last link in the support (Equation (6)). Starting with link 1 in free-flow, `bestNE` checks if the output of `freeFlowConfig` is a feasible assignment. If this is the case, the candidate assignment is the Best Nash Equilibrium, and `bestNE` terminates. If not, the free-flow link index is incremented by one, and the process is repeated until either a feasible assignment is found, or the number of links exceeds N , in which case no Nash equilibrium exists.

Algorithm 1 Best Nash Equilibrium

```

procedure bestNE( $N, r$ )
Inputs: Size of the network  $N$ , demand  $r$ 
Outputs: Assignment  $(x, m)$  at
             Best-Nash-Equilibrium

for  $i \in \{1, \dots, N\}$ :
    let  $(x, m) = \text{freeFlowConfig}(N, r, i)$ 
    if  $x_i \in [0, x_i^{\max}]$ :
        return  $(x, m)$ 
return No-Solution

procedure freeFlowConfig( $N, r, i$ )
Inputs: Size of the network  $N$ ,
           demand  $r$ , free-flow link index  $i$ 
Outputs: assignment  $(x, m) = (\bar{x}^{r,i}, \bar{m}^{r,i})$ 
           as defined in Eq's (5) and (6)

for  $i \in \{1, \dots, N\}$ :
    if  $i < j$ :
         $x_i = \hat{x}_i(j)$ ,  $m_i = 1$ 
    elseif  $i == j$ :
         $x_i = r - \sum_{k=1}^j x_k$ ,  $m_i = 0$ 
    else:
         $x_i = 0$ ,  $m_i = 0$ 
return  $(x, m)$ 

```

We first note from Algorithm 1 that from the definition of $\hat{x}_i(j)$, we can precompute $\hat{x}_i(j) \forall 1 \leq i < j \leq N$ in $O(N^2)$. The subroutine `freeFlowConfig` runs in $O(N)$ time. Finally, for each loop of the `bestNE` outer routine (with N iterations), the running time is a constant plus the running time of `freeFlowConfig`. Therefore, the overall running time of the algorithm is $O(N^2) + NO(N) = O(N^2)$.

V. INEFFICIENCY OF BEST NASH EQUILIBRIA

To study the inefficiency of Nash equilibria, in particular of the best Nash equilibrium, we use price of stability as a measure of inefficiency [1]. Price of stability is defined as the ratio between the cost of the best Nash Equilibrium and the social optimal cost. First we give an overview of social optimum for our model. Then we consider a simple two link parallel network and derive the price of stability for a triangular fundamental diagram. We show how this example illustrates the dependency of the price of stability on the flow demand and the free-flow latencies.

A. Social Optima

Consider an instance (N, r) where the flow demand r does not exceed the maximum capacity of the network $r \leq \sum_n x_n^{\max}$. Since the total cost function is $C(x, m) = \sum_{n=1}^N x_n l_n(x_n, m_n)$, the social optimum of the network is a solution to the optimization problem: $\min_{x, m} \sum_n x_n l_n(x_n, m_n)$ such that $\sum_n x_n = r$. It is shown in [7] that a socially optimal flow is necessarily in free flow on all links, which leads to an equivalent linear program: $\min_{x, m} \sum_n x_n a_n$ such that $\sum_n x_n = r$. Then, since the links are ordered by increasing free-flow latency $a_1 < \dots < a_N$, the social optimum is simply given by the assignment that saturates most efficient links first. Formally, if $k_0 = \max\{k | r > \sum_{n=1}^k x_n^{\max}\}$ then the social optimal assignment x^* is given by

$$x^* = (x_1^{\max}, \dots, x_{k_0}^{\max}, r - \sum_{n=1}^{k_0} x_n^{\max}, 0, \dots, 0) \quad (8)$$

B. Price of Stability on a Two-Link Network

To characterize the loss of efficiency of Nash equilibria several metrics have been used including price of anarchy [14] and price of stability [1]. The price of anarchy is defined as the ratio between the cost of the worst Nash equilibrium and the the social optimum cost, while the price of stability is defined as the ratio between the best Nash equilibrium and the social optimal cost. For the case of non-decreasing latency functions, the price of anarchy and the price of stability coincide since all Nash equilibria have the same cost by the essential uniqueness property [14]. Since we focus our analysis on the best Nash equilibrium, we use as a metric the price of stability.

Consider a network instance $(2, r)$ such that $a_1 < a_2$ and $x_2^{\max} + \hat{x}_1(2) > x_1^{\max}$. Let $\text{BNE}(2, r) = (x_{\text{BNE}}(r), m_{\text{BNE}}(r))$ be the best Nash equilibrium and $(x_{\text{SO}}(r), 0)$ be the social optimum, as defined by (8). The price of stability is then defined as $\text{POS}(N, r) = \frac{C(x_{\text{BNE}}(r), m_{\text{BNE}}(r))}{C(x_{\text{SO}}, 0)}$. From the expression (8) of the social optimum and the characterization of the best Nash equilibrium in Theorem 1, we only need to consider the following two cases:

a) *Case 1:* $0 \leq r \leq x_1^{\max}$: Using (8), all the demand will be on link 1 in free-flow. Similarly, from Theorem 1 we have that since link 1 can accommodate r in free-flow and the support cannot be smaller than a single link, then $\text{BNE}(2, r)$ has all flow demand on link 1 in free-flow, and

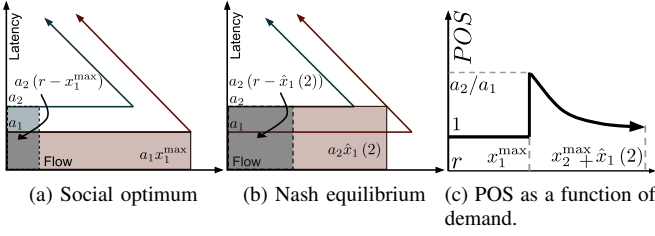


Fig. 4: Visualization of POS on two-link network. Differences in flow assignments between social optimum and Nash equilibrium are shown in 4a and 4b. The area of the shaded regions in 4a, 4b are the total costs attributed to each link. In 4c, the flat region corresponds to $r \leq x_1^{\max}$ (Case 1) and the decreasing region to $r > x_1^{\max}$ (Case 2).

is equivalent to the social optimum. In this case, the price of stability is equal to 1, i.e there is no decrease in efficiency due to selfish routing.

b) Case 2: $x_1^{\max} < r \leq x_2^{\max} + \hat{x}_1(2)$: We know that all flow demand cannot be accommodated by link 1. From (8), the social optimum assignment is given by $x_{\text{so}}(r) = (x_1^{\max}, r - x_1^{\max})$. From Theorem 1 we have that BNE $(2, r)$ has a single link in free-flow. Since the total demand exceeds the capacity of link 1, then under a best Nash equilibrium, link 2 is in free-flow, and link 1 is congested. Therefore $m_{\text{NE}}(r) = (1, 0)$. From Algorithm 1, the corresponding flow $x_{\text{NE}}(r)$ will be $(\hat{x}_1(2), r - \hat{x}_1(2))$. The comparison of the social optimum and Nash equilibrium assignments are depicted in Figure 4.

Computing the price of stability when $r > x_1^{\max}$ reveals where the inefficiencies lie in the Nash equilibrium. It is shown in [7] that $\text{POS}(2, r > x_1^{\max}) = \left(1 - \frac{x_1^{\max}}{r} \left(1 - \frac{a_1}{a_2}\right)\right)^{-1}$. In this simple two-link parallel network, the price of stability is maximal at $r = x_1^{\max}$ and equal to $\frac{a_2}{a_1}$ (Figure 4c). This shows in particular that for the general class of horizontal queueing congestion latencies, the price of stability is unbounded, since for any demand r and any positive constant A , we can design an instance $(2, r)$ such that the price of stability is $\frac{a_2}{a_1} > A$.

Also note that for a fixed flow demand $r > x_1^{\max}$, the price of stability is an increasing function of $\frac{a_2}{a_1}$. And as $a_2 \rightarrow a_1$, the price of stability goes to 1. Intuitively, the inefficiency of Nash equilibria can be directly attributed to the difference in free-flow latency between the links.

Additionally, as the demand $r \geq x_1^{\max}$ increases, the price of stability decreases. This occurs because the difference in total latency between social optimum and Nash equilibrium is constant for $r \geq x_1^{\max}$.

This also shows that selfish routing is most costly when a free-flow link is near maximum capacity (note the discontinuity in total latency for Nash equilibrium that occurs when demand exceeds the capacity of the first link $r > x_1^{\max}$). If a controller were to anticipate a scenario where demand was slightly above this capacity, they could dramatically reduce the inefficiency of the Nash equilibrium by rerouting a small fraction of the flow and keeping the link in free-flow.

VI. CONCLUSION

We introduced a new class of latency functions that models congestion in horizontal queueing networks, and we studied the resulting Nash equilibria for non-atomic congestion games on parallel networks. We showed that the essential uniqueness property does not hold for this new class, and that there may be up to $2N$ equilibria for a network instance (N, r) . Then we focused our attention on the best Nash equilibrium BNE (N, r) , which we proved is a single link free-flow equilibrium that minimizes the support, and then presented a constructive, quadratic time algorithm for finding this equilibrium. Finally, we derived price of stability results for an example network, then showed that if a link is anticipated to be near capacity, congestion can be completely averted by diverting only a small fraction of the demand.

REFERENCES

- [1] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, and T. Roughgarden. The Price of Stability for Network Design with Fair Cost Allocation. *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 295–304, 2004.
- [2] A. Aswani and C. Tomlin. Game-theoretic routing of GPS-assisted vehicles for energy efficiency. In *American Control Conference (ACC)*, 2011, pages 3375–3380. IEEE, 2011.
- [3] M. Babaioff, R. Kleinberg, and C. H. Papadimitriou. Congestion games with malicious players. *Games and Economic Behavior*, 67(1):22–35, 2009.
- [4] C. F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28(4):269–287, 1994.
- [5] T. L. Friesz and R. Mookherjee. Solving the dynamic network user equilibrium problem with state-dependent time shifts. *Transportation Research-Part B Methodological*, 40(3):207–229, 2006.
- [6] F. P. Kelly. Mathematical modelling of the Internet. *Mathematics unlimited-2001 and beyond*, pages 685–702, 2001.
- [7] W. Krichene, J. Reilly, S. Amin, and Bayen A. M. Nash equilibria and stackelberg routing on horizontal queueing networks, <https://sites.google.com/site/wkrichene/research/stackelberg>.
- [8] M. J. Lighthill and G. B. Whitham. On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178):317, 1955.
- [9] H. K. Lo and W. Y. Szeto. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research Part B: Methodological*, 36(5):421–443, 2002.
- [10] J. Morgan, K. Steiglitz, and G. Reis. The spite motive and equilibrium behavior in auctions. *The BE Journal of Economic Analysis & Policy*, 2(1):5, 2003.
- [11] C. Papadimitriou and G. Valiant. A new look at selfish routing. *Innovations in Computer Science (ICS)*, 2010.
- [12] M. Papageorgiou, J. M. Blosseville, and H. Hadj-Salem. Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. *Transportation Research Part B: Methodological*, 23(1):29–47, 1989.
- [13] T. Roughgarden. Stackelberg scheduling strategies. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 104–113. ACM, 2001.
- [14] T. Roughgarden and E. Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2):236–259, 2002.
- [15] D. Schmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4):295–300, 1973.
- [16] D. B. Work, S. Blandin, O. P. Tossavainen, B. Piccoli, and a. M. Bayen. A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress*, 2010(1):1, April 2010.
- [17] A. K. Ziliaskopoulos. A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation science*, 34(1):37, 2000.